EV Power - Lab 4 Project Report

Jingzhi Zhang

Example Solution 1

Part 0: libraries

```
#install.packages("gutenbergr")
#install.packages("tidytext")
#install.packages("maps")
```

```
library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────────── tidyverse 2.0.0
──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.1     ✔ stringr   1.5.2
✔ ggplot2   3.5.2     ✔ tibble    3.3.0
✔ lubridate 1.9.4     ✔ tidyr     1.3.1
✔ purrr     1.1.0
── Conflicts ──────────────────────────────────── tidyverse_conflicts()
──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```
library(gutenbergr)
library(tidytext)
library(dplyr)
library(stringr)
library(maps)
```

```
Attaching package: 'maps'

The following object is masked from 'package:purrr':

    map
```

```
renew_use_2023 <- read.csv("data/renew-use-2023.csv")
total_use_2023 <- read.csv("data/total-use-2023.csv")
view(renew_use_2023)
view(total_use_2023)
```

Part 1: Defining Research Question

Chosen Question: Which states had the highest proportion of renewable energy use in 2023?

Part 2: Data Preparation and Cleaning

```
total_use_2023_long <- total_use_2023 %>%
  pivot_longer(
    cols = -Energy_Source,
    names_to = "State",
    values_to = "total_use_2023"
  )
renew_use_2023 <- renew_use_2023 %>%
  mutate(
    State = str_to_upper(str_trim(State)),
    Renewable_Use_2023 = str_remove_all(
      Renewable_Use_2023,
      "kWh|MWh|GWh|[A-Za-z]|\\s"
    ),
    Renewable_Use_2023 = as.integer(Renewable_Use_2023)
  )

renew_use_2023 <- renew_use_2023 %>%
  select(State, Energy_Source, Renewable_Use_2023)

total_use_2023_long <- total_use_2023_long %>%
  mutate(
    Energy_Source = str_remove_all(Energy_Source, "\\(.*?\\)|†"),
    Energy_Source = str_trim(Energy_Source),
    Energy_Source = str_replace_all(Energy_Source, "-", "_"),
    Energy_Source = str_replace_all(Energy_Source, "\\s+", "_"),
    Energy_Source = str_to_lower(Energy_Source),
    State = str_to_upper(State),
    State = str_trim(State)
  )
view(renew_use_2023)
view(total_use_2023_long)
```

Part 3: Joining / Pivoting Datasets for Analysis

```
renew_state_2023 <- renew_use_2023 %>%
  group_by(State) %>%
  summarise(total_renewable_use_2023 = sum(Renewable_Use_2023, na.rm = TRUE))
```

```r
total_state_2023 <- total_use_2023_long %>%
  filter(
    Energy_Source == "total_renewable_energy" |
      Energy_Source == "total_energy_use"
  ) %>%
  group_by(State) %>%
  summarise(total_use_2023 = sum(total_use_2023, na.rm = TRUE))

merged_2023 <- renew_state_2023 %>%
  inner_join(total_state_2023, by = "State")

merged_2023 <- merged_2023 %>%
  mutate(
    renewable_share_percent = (total_renewable_use_2023 / total_use_2023) *
100
  )

view(merged_2023)
```

Part 4: Mapping Visualization

```r
us_states <- map_data("state")

state_lookup <- data.frame(
  state = tolower(state.name),
  abbr = state.abb
)

merged_map <- merged_2023 %>%
  left_join(state_lookup, by = c("State" = "abbr")) %>%
  left_join(us_states, by = c("state" = "region"))

ggplot(
  merged_map,
  aes(x = long, y = lat, group = group, fill = renewable_share_percent)
) +
  geom_polygon(color = "white", size = 0.2) +
  coord_fixed(1.3) +
  scale_fill_gradient(
    low = "lightyellow",
    high = "darkgreen",
    name = "Renewable Energy (%)",
    na.value = "grey90"
  ) +
  labs(
    title = "Share of Renewable Energy by State (2023)",
    subtitle = "Proportion of total energy use from renewable sources",
```
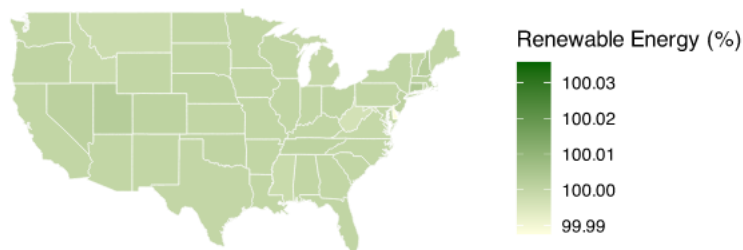
```
    caption = "Data: U.S. Energy datasets 2021–2023"
  ) +
  theme_minimal() +
  theme(
    legend.position = "right",
    plot.title = element_text(size = 16, face = "bold"),
    plot.subtitle = element_text(size = 12),
    axis.text = element_blank(),
    axis.title = element_blank(),
    panel.grid = element_blank()
  )
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
ℹ Please use `linewidth` instead.
```

Share of Renewable Energy by State (2023)

Proportion of total energy use from renewable sources

Data: U.S. Energy datasets 2021–2023